Chapter 12: Python Data Science Project

Contents

DRY

Don't **R**epeat **Y**ourself

Danger Will Robinson

This is going to be a weekly individual assignment ending with a Final Project due during finals week.

Always download the current version of this project.

Week 14: Get Started with Jupyter Notebooks

Time required: 120 minutes

Data Science Environments

Data science in Python can use the standard Python interpreter we have been using. Data scientists typically use IPython (Interactive Python) in Jupyter Notebooks. We will be using Google Colab.

Google Colab

Google Colab is a free cloud based Jupyter Notebook. You will need a Gmail account to use it.

1. Go to https://colab.research.google.com
2. Login with a Gmail account.

Python Data Science Primer

The Python Data Science Primer is where much of the Data Science work will be published.

1. Click the following link. https://github.com/itinstructor/JupyterNotebooks/blob/main/Notebooks/Python_Data_Science_Primer.ipynb
2. Click the Open in Colab button.

3. You may be asked to log into Google Colab with a Google Account.
4. The Notebook opens in read only mode. You can run any code cells, but you can't make changes.
5. You can go to File → Save a Copy in Drive. This will save an editable copy in your Google Drive.

NOTE: The Python Data Science Primer will be changing on a regular basis. Open a new copy each time you work on this project.

Colab Notebook Tutorials

The first is a tutorial for using Google Colab Notebooks.

- https://colab.research.google.com/notebooks/basic_features_overview.ipynb
- Markdown Guide for Google Colab (How to format a notebook) Notebook Markdown

Tutorial 1: Google Colab Python Data Science Learning Journal Notebook

1. Create a Google Colab notebook named: {Your Name} **Python Data Science Learning Journal**
2. For example: **Bill's Python Data Science Learning Journal**

Use this notebook to keep track of code and notes as you go through this project.

Use this notebook to store:

- Assignments, code, comments, and ideas about your understanding as we go.
- Think of your notebook as a Data Science learning journal.
- The quality of your notebook will be part of the grade for that week.

Tutorial 2: Python Data Science Primer – Python & Numpy

Please go through the following notebook to learn the basics of Python Data Science in a Jupyter Notebook. Keep track of what you are learning in your Learning Journal notebook.

NOTE: This notebook will be changing on a regular basis. Always open a new copy each time you work on assignments.

6. Click the following link.

https://github.com/itinstructor/JupyterNotebooks/blob/main/Notebooks/Python_Data_Science_Primer.ipynb

Tutorial 3: Numpy

As you go through these tutorials, put notes and code into your Learning Journal.

- https://www.w3schools.com/python/numpy/default.asp
- https://www.w3schools.com/python/numpy/numpy_intro.asp
- https://www.w3schools.com/python/numpy/numpy_getting_started.asp
- https://www.w3schools.com/python/numpy/numpy_creating_arrays.asp
- https://www.w3schools.com/python/numpy/numpy_array_indexing.asp
- https://www.w3schools.com/python/numpy/numpy_array_slicing.asp
- https://www.w3schools.com/python/numpy/numpy_array_shape.asp
- https://www.w3schools.com/python/numpy/numpy_array_reshape.asp
- https://www.w3schools.com/python/numpy/numpy_array_iterating.asp
- https://www.w3schools.com/python/numpy/numpy_array_sort.asp
- https://www.w3schools.com/python/numpy/numpy_random.asp

Reference: Python Data Science Handbook Chapter 2: Introduction to Numpy

Assignment Submission

Each section of the notebook has exercises.

- Follow the directions in the Notebook for the exercises.
- Insert screenshots of your Google Collab Notebook Exercises solution successful run.

- Insert a Share Link for your Google Collab Notebook.
- Submit in Blackboard.

Week 15

Time required: 120 minutes

Tutorial 4: Pandas Tutorials

As you go through these tutorials, put notes and code into your Learning Journal.

- https://www.w3schools.com/python/pandas/pandas_intro.asp
- https://www.w3schools.com/python/pandas/pandas_getting_started.asp
- https://www.w3schools.com/python/pandas/pandas_series.asp
- https://www.w3schools.com/python/pandas/pandas_dataframes.asp
- https://www.w3schools.com/python/pandas/pandas_csv.asp
- https://www.w3schools.com/python/pandas/pandas_json.asp
- https://www.w3schools.com/python/pandas/pandas_analyzing.asp
- https://www.w3schools.com/python/pandas/pandas_cleaning.asp
- https://www.w3schools.com/python/pandas/pandas_cleaning_empty_cells.asp
- https://www.w3schools.com/python/pandas/pandas_cleaning_wrong_format.asp
- https://www.w3schools.com/python/pandas/pandas_cleaning_wrong_data.asp
- https://www.w3schools.com/python/pandas/pandas_cleaning_duplicates.asp

Reference: Python Data Science Handbook Chapter 3: Data Manipulation with Pandas

Tutorial 5: Python Data Science Primer – Pandas

Please go through the following notebook to learn the basics of Python Data Science in a Jupyter Notebook. Keep track of what you are learning in your Learning Journal notebook.

NOTE: This notebook will be changing on a regular basis. Always open a new copy each time you work on assignments.

https://github.com/itinstructor/JupyterNotebooks/blob/main/Notebooks/Python_Data_Science_Primer.ipynb

Assignment Submission

Each section of the notebook has exercises.

- Follow the directions in the Data Primer Notebook for the exercises.
- Insert screenshots of your Google Collab Notebook Exercises solution successful run.
- Insert a Share Link for your Google Collab Notebook.
- Submit in Blackboard.

Week 16

Time required: 120 minutes

Tutorial 6: Python Data Science Primer – Matplotlib

Please go through the following notebook to learn the basics of Python Data Science in a Jupyter Notebook. Keep track of what you are learning in your Learning Journal notebook.

NOTE: This notebook will be changing on a regular basis. Always open a new copy each time you work on assignments.

https://github.com/itinstructor/JupyterNotebooks/blob/main/Notebooks/Python_Data_Science_Primer.ipynb

Tutorial 7: Matplotlib Tutorials

As you go through these tutorials, put notes and code into your Learning Journal.

1. https://www.w3schools.com/python/matplotlib_intro.asp
2. https://www.w3schools.com/python/matplotlib_getting_started.asp
3. https://www.w3schools.com/python/matplotlib_pyplot.asp
4. https://www.w3schools.com/python/matplotlib_plotting.asp

Reference: Python Data Science Handbook Chapter 5: Visualization with Matplotlib

Tutorial 8: Machine Learning Tutorials

As you go through these tutorials, put notes and code into your Learning Journal.

1. https://www.w3schools.com/python/python_ml_getting_started.asp
2. https://www.w3schools.com/python/python_ml_mean_median_mode.asp

Tutorial 9: Python Data Science Primer – Pandas COVID Analysis Project

Please go through this tutorial in your Learning Journal. This will give you a start on how to create your final project.

NOTE: This notebook will be changing on a regular basis. Always open a new copy each time you work on assignments.

https://github.com/itinstructor/JupyterNotebooks/blob/main/Notebooks/Python_Data_Science_Primer.ipynb

Assignment: Rough Draft Final Data Analysis Project

Thinking about what type of data project you want to work on. There are thousands of datasets that can be used. This can be done individually or as pair programming.

Data Science is turning raw data into information that has meaning.

Start gathering information and laying out your final project.

Create a separate Colab notebook for this project.

1. Project
    a. Describe your project idea.
    b. What question are you trying to answer?
2. Dataset
    a. What dataset do you think you will use?
3. Analysis
    a. How will you analyze the data?
4. Conclusions
    a. This can't be filled in until your project is complete.

Assignment Submission

Each section of the notebook has exercises.

- Submit a draft of your Data Analysis project. This can be screenshots in Word or a link to your Colab notebook
- Follow the directions in the Data Primer Notebook for the exercises.

- Insert screenshots of your Google Collab Notebook Exercises solution successful run.
- Insert a Share Link for your Google Collab Notebook.
- Submit in Blackboard.

Final Project: Data Analysis

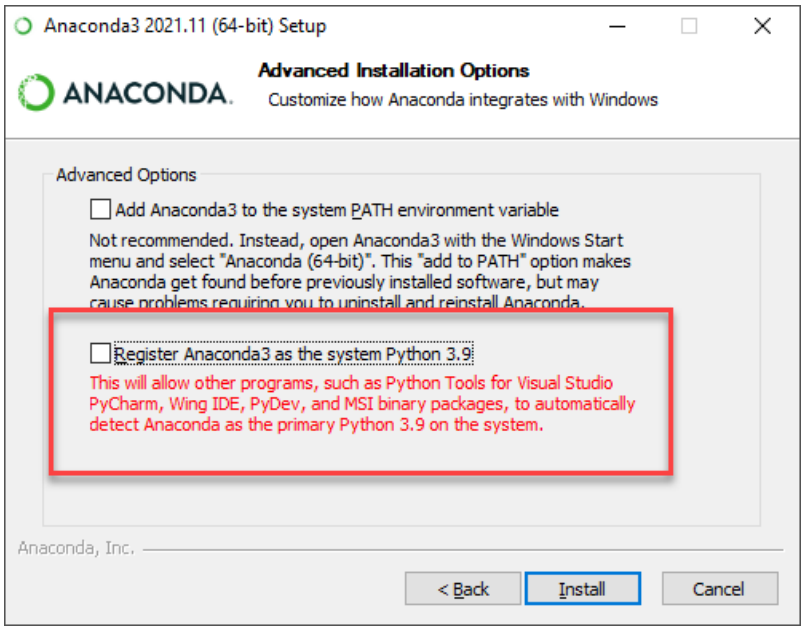
Assignment Submission

- Follow the directions in the Data Science Primer Notebook.
- Insert screenshots of your Google Collab Notebook.
- Insert a Share Link for your Google Collab Notebook.
- Submit in Blackboard.

Optional: Setup Anaconda

Anaconda is a local environment for Data Science and Jupyter Notebooks.

1. Go to https://www.anaconda.com/ Download the version for your operating system.
2. Double Click the installation file.
3. Click Next, **Install for everyone**.
4. **Uncheck** Register Anaconda3 as the system Python. This can cause trouble with other Python installations.



5. Click Install.
6. Click Start → Type **Anaconda Navigator**

Open Jupyter Notebook on Other Drives

By default, Jupyter Notebook opens in your profile folder.

1. Start → **Anaconda Prompt**
2. Enter: **jupyter notebook --notebook-dir=/E:**
   Change E: to whatever drive you wish to work in.

Update Anaconda

Anaconda Prompt → Right Click → More → Run as Administrator

conda update --all